

Random sequencing of cDNAs and identification of mRNAs

James V. Anderson

Corresponding author. U.S. Department of Agriculture, Agricultural Research Service, Biosciences Research Laboratory, Plant Science Research, 1605 Albrecht Boulevard, P.O. Box 5674, Fargo, ND 58105; andersjv@fargo.ars.usda.gov

David P. Horvath

U.S. Department of Agriculture, Agricultural Research Service, Biosciences Research Laboratory, Plant Science Research, 1605 Albrecht Boulevard, P.O. Box 5674, Fargo, ND 58105

As a first step toward developing a genomics-based research program to study growth and development of underground adventitious shoot buds of leafy spurge, we initiated a leafy spurge expressed sequence tag (EST) database. From the approximately 2,000 clones randomly isolated from a cDNA library made from a population containing growth-induced underground adventitious shoot buds, we have obtained ESTs for 1,105 cDNAs. Approximately 29% of the leafy spurge EST database consists of expressed genes of unknown identity (hypothetical proteins), and 10% represents ribosomal proteins. The remaining 60% of the database is composed of expressed genes that show BLASTX sequence identity scores of ≥ 80 with known GenBank accessions. Clones showing sequence identity to a *Histone H3*, a gibberellic acid-responsive gene, *Tubulin*, and a light-harvesting chlorophyll a/b-binding protein were shown to be differentially expressed in underground adventitious shoot buds of leafy spurge after breaking of dormancy. RNA encoding a putative cyclin-dependent protein kinase (CDK)-activating kinase, a gene associated with cell division, and *Scarecrow-like 7*, a gene involved in GA signaling, were present at similar levels in dormant and growth-induced underground adventitious shoot buds. These data show how even a small EST database can be used to develop a genomics-based research program that will help us identify genes responsive to or involved in the mechanisms controlling underground adventitious shoot bud growth and development.

Nomenclature: Amp, ampicillin; CAK, CDK-activating protein kinase; CDK, cyclin-dependent protein kinase; EST, expressed sequence tag; LB, Luria–Bertani broth; Lhcb, light-harvesting chlorophyll a/b-binding protein; NCBI, National Center for Biotechnology Information; SSC, sodium chloride sodium citrate solution.

Key words: Cell cycle, dormancy, genomics, perennial weeds, signal transduction.

In 1980, published work estimated that $\sim 75,000$ different genes were present in the average plant genome and that $\sim 27,000$ different genes are expressed in leaf tissue (Kamlay and Goldberg 1980). However, recent advances in sequencing now indicate that the genomes of *Arabidopsis* and rice (*Oryza sativa* L.) contain fewer than 26,000 different genes (The Arabidopsis Genome Initiative 2000), with about 35,000 and 6,000 different genes present in the genomes of human and yeast, respectively (Pennisi 2001). Thus, it is possible that considerably fewer genes are expressed in any given tissue than previously thought. With the increasing number of sequenced and characterized genes present in public databases (National Center for Biotechnology Information [NCBI], SWISS-PROT, EMBL, etc.), it may be possible to identify important and useful genes by comparing “single-pass” sequence data (also known as expressed sequence tags, ESTs) obtained from the sequencing of randomly isolated clones from any given cDNA library. We have undertaken the sequencing of random cDNA clones from underground adventitious buds (hereafter referred to as buds) of leafy spurge (*Euphorbia esula* L.) to determine if this approach would be useful for advancing our knowledge on signaling pathways that control the growth and development of vegetative buds of perennial weeds.

Leafy spurge is an invasive, deep-rooted, perennial weed that propagates vegetatively from buds located on the subterranean portion of the stem (crown) and lateral roots (Coupland et al. 1955). Leafy spurge infests more than 5

million acres of land in 36 states and in the prairie provinces of Canada and is responsible for an estimated annual loss of \$130 million in the four-state region of the Dakotas, Montana, and Wyoming alone (Leitch et al. 1996). In fact, leafy spurge causes such a threat to native vegetation in pastures, rangelands, and native habitats (Bangsund et al. 1999; Biesboer and Eckardt 1996) that the Nature Conservancy has termed leafy spurge as “one of the dirty dozen of America’s least wanted invasive species of U.S. ecosystems” (Stein and Flack 1996). Because of its vegetative reproduction, leafy spurge is resistant to many control methods (Sell et al. 1998). Each crown or root bud can regenerate a new plant after treatment of foliage with herbicides, biological control agents, mowing, or grazing with sheep. Recent observations indicate that even after 7 to 8 yr of control with flea beetles (one of the most successful biological control agents to date), 10 to 15 root buds per 10 cm² of topsoil remain on the root system of leafy spurge plants (Kirby et al. 2000). Current assessments estimate that by the year 2025, biological controls may reduce leafy spurge by 65% or decrease the economic effect by \$58.4 million (Bangsund et al. 1999). However, this would still leave an estimated 0.7 to 1 million acres of infested land in the northern great plains area alone.

In leafy spurge, auxin and leaf-derived signals other than auxin (most likely sugar) control correlative inhibition of buds (Horvath 1999). Removal of aboveground foliar tissue releases buds from correlative inhibition within the first 24 to 48 h as determined by increased levels of cell cycle activity

(Horvath and Anderson 2000). These data suggest that correlative inhibition of buds may be due to a blockage of cell division or interaction with signaling pathways controlling the cell cycle. To identify differentially expressed genes for the study of signaling pathways associated with bud growth and development, we have initiated a genomics-based approach to study developmental regulation in buds including single-pass sequencing and cataloging of random cDNA clones from a cDNA library for the development of an EST database. Our genomics-based approach will eventually incorporate DNA arrays developed from an EST database for genes expressed during the growth and development of leafy spurge buds.

The development of EST databases to enhance genomic studies in plants has been used since the early 1990s (Keith et al. 1993; Park et al. 1993; Uchimiya et al. 1992) and is becoming commonplace in research programs; however, very little attention has been directed toward developing an EST database for perennial weeds that could be used in a genomics-based research program. This report describes the first 1,105 ESTs obtained from a cDNA library constructed from leafy spurge buds. To show the immediate effect of this research project, the expressions of several key genes identified from the leafy spurge EST database were further characterized.

Because the variable growth and development of buds allow leafy spurge to escape most control measures, new knowledge gained from genomics-based research will be essential for identification of biochemical and signal transduction pathways and potential sites of action for manipulation or control of vegetative reproduction of this weed. Indeed, the economic and environmental catastrophe caused to ranchers, land managers, and taxpayers in the United States and Canada alone has justified the need for new knowledge to improve integrated pest management systems to control leafy spurge.

Materials and Methods

Plant Material

Plants used for these experiments were started from shoot cuttings collected from a greenhouse population that originated from a small group of plants isolated from a wild leafy spurge population in North Dakota. Shoot cuttings were stripped of the lower 5 cm of leaves, dipped in Rootone®,¹ placed in Sunshine Mix 1,² and grown in 25- by 203-mm low-density SC-10 Super Cell Ray Leach Cone-tainers®³ in a greenhouse under an 18-h photoperiod at approximately 28 ± 4 C for 2 to 3 mo. Plants were watered with tap water and fertilized twice weekly with Peters® 20–20–20 (N-P-K) fertilizer. Plants used for all of the studies below were single stems with 70 to 100 leaves and an average of 56 buds per plant (SD = 20). All distinguishable root buds (those below the crown and 0.25 mm or larger) were harvested.

cDNA Library Construction

Buds (shoot buds below the crown) of leafy spurge were harvested 3 d after excision of the aerial portion (including the removal of crown buds) of the plant. To prepare the amount of RNA needed for the library, buds from approximately 10 sets of 21 individual plants were collected over

several months and pooled. Previous studies have indicated that this represents a mixed population of dormant and growing buds with approximately 60% of the buds incapable of growth (endo-dormant) (D. P. Horvath, unpublished data). Total RNA was harvested from the buds according to previously published methods (Horvath and Olson 1998). Ten micrograms of messenger RNA was isolated from approximately 10 mg of total RNA using the Poly ATract mRNA Isolation System III.⁴ Messenger RNA was transferred to Stratagene Corp. (La Jolla, CA) for construction of the custom cDNA library. Resulting cDNAs were directionally ligated into the Hybri-Zap 2.1 XR two-hybrid vector.⁵ The resulting library had over 3 million independent inserts with an approximate average size ($n > 500$) of 1.5 kbp. Mass excision and plasmid rescue of the library was accomplished using protocols and reagents supplied by Stratagene. The resulting plasmid library was transformed into *E. coli* strain DH5-alpha for all further manipulations.

Plasmid Template Preparation

Individual colonies were randomly picked from a Luria-Bertani broth (LB)/ampicillin (Amp) plate and transferred to 10 ml of LB containing 80 µg ml⁻¹ Amp and incubated overnight at 37 C at 200 rpm. The following day, the overnight cultures were centrifuged at 2,000 × *g*, and the resulting bacterial pellets were each resuspended into 400 µl of freshly prepared buffer (25 mM Tris [pH 8.0], 50 mM glucose, and 10 mM EDTA). Plasmid DNA was isolated from 200 µl of each resuspended culture following the Modified Alkaline-Lysis/PEG Precipitation Procedure outlined in the Automated DNA Sequencing Chemistry Guide (Document No. 4305080, PE Applied Biosystems [available from: www.pebiiodocs.com]). The remaining 200 µl of resuspended bacterial pellet was used to make glycerol stock solutions. cDNA insert sizes were determined by incubating 2 µl of each purified plasmid with a combination of *EcoRI* and *XhoI* restriction enzymes for 1 h at 37 C followed by separation on a 1% agarose gel.

DNA Sequencing Protocol

Inserts from plasmids were sequenced using the ABI Prism® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit Protocol.⁶ Each sequencing reaction was made 1 × by adding 4 µl of Ready Reaction mix, 4 µl of 2.5 × sequence buffer (200 mM Tris [pH 9.0] and 5 mM MgCl₂), 1 µl of template (1 µg/µl purified plasmid), 1 µl of 5 pmol/µl pAD5-primer [5'-AAAGAGATCGAATTAGGATC-3'], and 10 µl of sterile H₂O to equal a final reaction volume of 20 µl. Sequence reactions were performed using a PTC-200 Peltier Thermal Cycler.⁷ The program conditions were one cycle of 3 min at 95 C, followed by 30 cycles of 45 s at 98 C, 10 s at 45 C, and 4 min at 60 C. At the end of the sequencing runs, each sequence reaction was desalted using a DyeEx spin column⁸ according to the manufacturer's protocol. The desalted sequence reactions were then dried under vacuum and sent to the DNA Sequencing Facility at Iowa State University for automated sequencing.

RNA Extraction and Northern Blotting

Buds from 14 to 21 individual plants (per time point) from the greenhouse population were harvested, pooled, and

immediately frozen in liquid nitrogen. Buds from untreated plants were harvested and assumed to be equivalent to 0 h after treatment (control), as it is unlikely that any change in gene expression, detectable by northern analysis, would occur during the few minutes of harvesting time. Pooled buds were ground to a fine powder with a mortar and pestle and stored at -80°C for later use. RNA was extracted using a previously described extraction method (Chang et al. 1993). Total RNA was quantified using a GeneQuant⁹ spectrophotometer. Total RNA ($50\text{ }\mu\text{g lane}^{-1}$) was separated on a 1% denaturing agarose gel and blotted onto a positively charged nylon membrane using standard protocols (Sambrook et al. 1989). Ethidium bromide staining intensity of ribosomal RNA bands was quantified in gels using a Fluor-S MultiImager¹⁰ and was the method used to determine consistent loading of RNA for each time point. Gel-purified cDNA inserts were radiolabeled by nick translation and hybridized overnight at 42°C (50% formamide solution) to immobilized RNA on the filter using standard protocols (Sambrook et al. 1989). Filters were washed four times for 5 min each in $2\times$ sodium chloride sodium citrate solution (SSC), 0.1% (v/v) sodium dodecyl sulfate (SDS) at room temperature, followed by two washes in $2\times$ SSC at 65°C for 15 min each (solution formulas for SSC and SDS are described in Sambrook et al. 1989). Radioactive signals were visualized and quantified using a Packard Instant Imager¹¹ with approximately a 1-h exposure. Following visualization, filters were submerged once in boiling 0.1% (v/v) SDS solution and allowed to cool to room temperature. Filters were rinsed with 0.1% (v/v) SDS solution at room temperature, and removal of all radioactivity was ensured by visualization of the clean filter for 1 h on the imager prior to reprobing with a new cDNA. Time course experiments were repeated twice with independently isolated sets of RNA for each time point.

Sequence Analysis and EST Database Development

Sequence trace files (chromatograms) from each sequence reaction were edited to remove 5' or 3' vector sequence, or any poor-quality sequence at the end of trace files, using the Lasergene 99 software program.¹² Each edited EST sequence (average size ~ 500 bp) was used in a GenBank BLASTX search (Altschul et al. 1997) using the NCBI World Wide Web site (<http://www.ncbi.nlm.nih.gov/BLAST/>) to obtain the best match sequence identities. Scores of ≥ 80 (minimum window size of 50 to 100 amino acids) were used as a benchmark for indicating potentially significant homology (Newman et al. 1994). Sequence identities < 80 were considered to be unknowns. Open reading frames for unknowns due to low-identity match scores were confirmed using Lasergene 99 software. The first 468 ESTs obtained from our cDNA library were submitted to GenBank's EST database (dbEST; Boguski et al. 1993), where they were given GenBank accession numbers and kept in a publicly available archive (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>).

Results and Discussion

Prior to initiating our genomics-based approach, only six genes from leafy spurge had been entered into the Genbank database. The lack of genetic resources for perennial weeds such as leafy spurge made it difficult to initiate a genomics-

based research program to study the growth and development of buds. Thus, our first objective was to develop an EST database for genes expressed in buds of leafy spurge. From our existing 2,000 isolated cDNAs, we report here EST data for 1,105 clones and accession numbers for the first 468 clones sequenced. The leafy spurge ESTs that have accession numbers can be obtained from dbEST (see Materials and Methods). The current list of BLASTX results for each EST in the database can be accessed and viewed through the Center for Computational Genomics and Bioinformatics, University of Minnesota (<http://web.ahc.umn.edu/biodata/euphorbia/>). Table 1 shows the putative protein identity match and GenBank accession numbers for each of the first 265 ESTs from the original 468 ESTs that showed protein identity matches other than ribosomal proteins or unknowns. Interestingly, the EST database for leafy spurge is represented by greater than 29% unknowns (unknowns include ESTs with matches to hypothetical proteins, no matches, or matches with BLASTX scores less than 80). The percentage of unknowns identified in our database closely parallels the 30% of unknowns recently reported for strawberry (Aharoni et al. 2000). Of the 29.8% (329 ESTs) of unknowns identified in our EST database, 222 were matches to hypothetical proteins, and the remaining 107 either gave no BLASTX identity matches or had scores of less than 80. Prior to assigning the previously unidentified leafy spurge genes (ESTs) as unknowns (hypothetical proteins), a high-quality open reading frame was identified (greater than 100 amino acids without a stop signal). Leafy spurge ESTs that had the greatest identity to unknown (hypothetical) proteins and that had BLASTX scores of ≥ 80 most often showed sequence identity to unknowns previously identified from *Arabidopsis thaliana*. At 10%, ribosomal proteins make up the next greatest percentage of the leafy spurge EST database. Other categories of identified genes that constitute at least 1% or greater of the 1,105 ESTs in our leafy spurge EST database are listed in Table 2. Overall, the EST database contains approximately 20% redundancy. Additionally, some of the clones isolated for the EST database have been fully sequenced and further characterized to provide us with information on the growth and development of buds (Anderson and Horvath 2000; Chao et al. 2000).

Differentially Expressed Genes Identified by Sequence Analysis

As part of our genomics-based research program, we are interested in identifying genes responding to signals that control bud development in perennial plants and weedy species. Homology searches of sequences in our leafy spurge EST database revealed a number of genes that were likely to be differentially expressed concomitantly with resumed shoot bud growth. These genes included several with homology to cell division genes (*Histone H3*, *Tubulin*, and cyclin-dependent protein kinase [CDK]-activating kinase [CAK]), a gene with similarity to a GA-responsive gene (*GASA*) from *A. thaliana*, a gene with homology to a gene suspected to play a role in GA signaling (*Scarecrow-like 7*), and a gene with similarity to light-harvesting chlorophyll *a/b*-binding protein (*Lhcb1*) that is required for development of the photosynthetic apparatus (Anderson and Horvath

TABLE 1. List of putative protein identity matches for expressed sequence tags (ESTs) obtained from randomly isolated cDNAs developed from *Euphorbia esula* underground adventitious root buds. GenBank accession numbers are shown for each submitted EST.

Putative protein	Accession No.	Putative protein	Accession No.	Putative protein	Accession No.
ABC transporter	AW821911	Chlorophyll a/b-binding protein (LHCII Type 1)	AW832679	Elongation Factor 1-beta	AW862629
Acetyltransferase	AW821912			Elongation Factor 1B gamma	BE123868
Actin	AW821913	Chlorophyll a/b-binding protein (LHCII Type 1)	AW832680	Elongation factor (EF-2)	AW862630
Actin	AW821914			En/Spm-like transposon	AW862641
Actin Depolymerizing Factor 1	AW821915	COP1 interacting protein	AW832681	Endoxyloglucan transferase	AW840615
Acyl-CoA synthetase	AW821916	COP1 regulatory protein	AW832682	Endoxyloglucan transferase	AW840607
Adenosylhomocysteinase	AW821917	Cryptochrome 2 apoprotein	AW832683	Enolase (2-phosphoglycerate dehydratase)	AW944678
ADP/ATP Carrier Protein 1 precursor	AW821918	Cyclic nucleotide-calmodulin-regulated ion channel	AW832684	Epoxide hydrolase	AW862631
Alcohol dehydrogenase	AW821919			Esterase D	AF227624
Aldehyde dehydrogenase	AW821920	Cyclin-selective ubiquitin-carrier protein	AW990944	Ethylene inducible	AW862616
Aminolevulinatase dehydratase	AW821921	Cyclophilin	AF242312	Eukaryotic Translation Initiation Factor 3, Subunit 8	AW990936
Aminopeptidase	AW821922	Cyclophilin	AW832685		
AMP binding protein	BE056345	Cytochrome P450	AW832686	Eukaryotic Translation Initiation Factor 3, Subunit 8	AW990937
AP2 domain containing	AW821923	Cytochrome P450	BE231327		
Aquaporin (plasma membrane)	AW990929	Cytochrome P450	BE231346	Eukaryotic Initiation Factor 4B	AW862632
Aquaporin (plasma membrane)	BE231342	Dehydration responsive	AW840608	Expansion precursor	AW862617
Aquaporin (plasma membrane)	AW821924	Dehydroquinase shikimate dehydrogenase	AW840595	Farnesyl pyrophosphate synthetase	AW862618
Aquaporin (tonoplast)	AW821925	Disease resistance protein putative	AW840610	Fatty acid oxidation tetrafunctional protein precursor	AW990925
Aquaporin (tonoplast)	AW990927	Desiccation protectant protein (LEA14)	AF239929		
Aquaporin (tonoplast)	AW990928	Dihydrolipoamide dehydrogenase	AW840596	Ferridoxin-thioredoxin reductase	AW862619
Aquaporin (tonoplast)	AW990930	DNA-binding protein	AW840598	Flavone synthase	AF228663
Argonaute	AW821926	DNA-binding protein/ethylene inducible	AW840599	Formate-tetrahydrofolate ligase	AW862638
ARP-1 N-acetyltransferase	AW821927	DNA-binding protein/ethylene responsive	AW840611	Fructose-diphosphate aldolase	AW862620
ATPase beta subunit	AW821928	DNA-binding protein/transcription factor AP2	AW840600	Fructose-diphosphate aldolase	AW862621
ATP citrate lyase	AW821929			Fructose-bisphosphate aldolase putative	BE095276
ATP-dependent RNA helicase	AW944687	DNA-binding protein/WRKY3 transcription factor	AW840601	G-Protein-coupled receptor	AW862633
ATP phosphorybosyl transferase	AW821930			GDSL-motif lipase/hydrolase	AW862622
ATP synthase beta chain, mitochondrial precursor	BE231338	DNA-damage/tolerance protein	AW840597	Gibberellin-regulated protein	AW862634
Auxin down-regulated (ARG10)	AW821931	DnaJ	AF239932	Gibberellin-regulated protein	AW862635
BAP3-like	AW832661	DnaJ	AW840612	Glutamine synthase cytosolic isozyme	AW862636
Beta-hydroxyacyl-ACP dehydratase	AW832662	DnaJ	AW840602	Glutathione peroxidase	AW862623
Calcium pump/ER type	AW832665	DnaJ	AW840603	Glutathione S-Transferase, auxin inducible	AF239928
Calmodulin	AW832666	DnaJ	AW840613		
Carbonic anhydrase, chloroplast precursor	AW832667	DnaJ	AW840614	Glutathione S-Transferase, Phi Class	AF242309
Carbonyl reductase	AW832668	DnaK	BE231328	Glutathione S-Transferase, Phi Class	AW862639
Catalase	AW832669	DnaK	AW840604	Glutathione S-Transferase, Theta Class	AF239927
Catalase	AW832670	DnaK-type Molecular Chaperone HSP70	AW840605	Glutathione S-Transferase, Theta Class	AF263737
Cathepsin B cysteine proteinase	AW832671	Dynamitin-like protein	BE231348	Glycine-rich RNA-binding protein	AW862624
CDK-activating kinase	AF230740	Elongation Factor 1-alpha	AW840606	Glycoprotein EP1	AW874978
Cell elongation protein diminuto	AW832672	Elongation Factor 1-alpha	AW862625	GTPase activator protein (RAB-like)	AW874979
Cellulose synthase-catalytic chain subunit	AW832673	Elongation Factor 1-alpha	AW862640	GTP-binding protein	AW874980
Centromere protein	AW832674	Elongation Factor 1-alpha	AW862637	GTP-binding protein	BE231339
CGI-like protein	AW832675	Elongation Factor 1-alpha	AW862626	GTP-binding protein/RAB1	AW874981
Chalcone-flavanone isomerase	AW832676	Elongation Factor 1-alpha	AW862627	GTP-binding protein/RAB6	AW874982
Chlorophyll a/b-binding protein	AF220527	Elongation Factor 1-alpha	AW862613	Guanine nucleotide-binding protein beta subunit	AW874983
Chlorophyll a/b-binding protein (LHCI)	AW832677	Elongation Factor 1-alpha	AW862614		
Chlorophyll a/b-binding protein (LHCI)	AW832678	Elongation Factor 1-alpha	AW862628	H+ transporting ATP synthase	AW874984
		Elongation Factor 1-alpha	AW862615	HR-like lesions able to induce	AW874985

TABLE 1. Continued.

Putative protein	Accession No.	Putative protein	Accession No.	Putative protein	Accession No.
Heat Shock Protein 70-Cytosolic	AW874986	Nascent polypeptide-associated complex	AW875016	RNA-binding protein	AW944686
Heat Shock Protein 70-Cofactor (GRPE protein)	AW874987	alpha chain		RNA helicase	AW944688
Heat Shock Protein 80	AF221856	Nascent polypeptide associated complex	AW875017	RNA Polymerase I, II, and III 24-kDa subunit	AW944689
Heat shock protein (HSP81-1)	BE231337	Nucleotide sugar epimerase	AW875018		
Heat Shock Protein 90 (GRP94)	AF239931	Nucleoid DNA-binding protein	AW875019	Rubisco, Small Subunit N-methyltransferase I	AW944690
Heat Shock Protein Precursor-(22 kDa)-Mitochondrial	AF237957	Nucleoside diphosphate kinase	AW875020	Rubisco, Small Subunit N-methyltransferase I	AW944691
Heat shock transcription factor (HSF8)	AW874988	Omega 6 fatty acid desaturase	AW875021		
Hemolysin	BE231343	P-type transporting ATPase	AW875022	S-Adenosylmethionine decarboxylase proenzyme	AW944692
Histone acetyltransferase	AW874990	Pectinesterase precursor	AW875023		
Histone acetyltransferase	AW874991	Pectinesterase precursor	AW875024	S-Adenosylmethionine Synthetase 2	AW944693
Histone H1	AF222804	Peroxidase	AW875025	SAH7 protein (allergen-like)	BE231353
Histone H1	AW874992	Peroxidase	AW875026	Scarecrow (gibberellin responsive modulation protein)	AW944694
Histone H1-like protein	AW874993	Peroxidase	AW875030		
Histone H2A	AF242311	Peroxidase	AW875027	Serine/threonine protein kinase	AW944695
Histone H2A	AW874994	Phosphate-inducible protein	BE231329	Serine/threonine protein kinase	AW944697
Histone H2B	BE231352	Phosphate-inducible protein	AW875028	Serine/threonine protein kinase (similar to WPK4)	AW944698
Histone H3	AF239930	Phosphoglycerate dehydrogenase precursor	AW875029		
Histone H3	AW874995	Phospholipase D precursor	AW944677	Shaggy-like protein kinase	AW944699
Histone H3.2	AW874996	Phytochelatin synthetase-like protein	AW944679	Similar to Arabidopsis Clone F16A14.8	BE231335
Histone H3.2	AW874997	PLTSLRE-type protein kinase	BE231351	Similar to Arabidopsis Clone T7N9.15	BE231341
Imbibition protein	AW874998	Polyphosphoinositol/Phosphatidylcholine transfer protein	AW821910	Similar to Arabidopsis Developmental Protein DG1118	AW944700
Imbibition protein	AW874999		AW944680		
Iron-regulated transporter	AW875000	Polyubiquitin	AF221858	Similar to gene product from drosophila	AW944701
Isocitrate dehydrogenase	AW875001	Polyubiquitin	AW944681	Similar to Pfam family	AW944702
Isocitrate dehydrogenase	AW875002	POP3	AW944682	Similar to Shock Protein SRC2 from soybean	AW944703
Lipid transferase-nonspecific	AW875003	Pore Protein 24 K chain	AW944683		
Lipid transferase precursor-nonspecific	BE231326	Potassium channel beta subunit	AW944706	Skp1 (putative kinetochore protein)	AW944704
Low-affinity calcium antitransporter	AW875004	Prohibitin	AW944707	Small nuclear ribonucleoprotein-associated protein	BE231330
Magnesium Chelatase Subunit CHLD	AW875005	Protein disulfide isomerase	AW944709		
Magnesium-Protoporphyrin IX Methyltransferase	AW875006	Protein disulfide isomerase	AW944710	Steroid-binding protein	BE231347
Malate dehydrogenase	AW875007	Protein phosphatase 2C	AW944711	Stomatin-like protein	BE060022
Malate dehydrogenase	AW875008	Proteosome 27-kDa subunit	AF227625	Stress-related protein	AW944705
MAP kinase (MAPk)	AF242308	Proteosome Component C3 (Macropain Subunit e3)	AW944713	Subtilisin-like proteanase	AW944708
MAP kinase (MAPkk)	BE231345	Pyruvate dehydrogenase, beta subunit	AW840609	Succinyl-CoA ligase, beta subunit	AW990920
Metallothionein-like	AW875009	Pyruvate kinase-cytosolic	AW944714	Succinyl-CoA ligase, beta subunit	AW990921
Metallothionein-like	BE231333	Raffinose synthetase	AW944715	Sucrose synthase	AW990923
Monosaccharide transport protein	AW875010	Receptor-like protein kinase	AW944716	Sucrose transport protein	AF242307
MtN3-like protein	AW875011	Receptor-like protein kinase	AW944717	Super oxide dismutase-Mn	AF242310
Multicatalytic Endopeptidase Complex Subunit 5	AW875012	Receptor-like protein kinase	AW944718	T-complex protein, epsilon subunit	AW990924
NADH-ubiquinone oxidoreductase	AW875014	Receptor-like protein kinase	AW944684	Thaumatococcus	AW990926
NADPH-protochlorophyllide oxidoreductase	AW875015	Retinoblastoma	AW944696	Thioredoxin-like protein	AW990952
		Retinoblastoma	AF230739	Tic20 chloroplast protein of import apparatus	AF227619
		Ring zinc-finger protein	AW944685	Transmembrane transporter protein	AW875013
			BE095285	Transport Inhibitor Response 1	AW990931

TABLE 1. Continued.

Putative protein	Accession No.	Putative protein	Accession No.	Putative protein	Accession No.
Transport Inhibitor Response 1	AW990932	Tubulin-beta chain	BE231350	1-acyl-sn-glycerol-3-phosphate acyltransferase	BE095286
Transport protein (SEC12p)	AW990933	Tubulin-beta	AW832663	1-Deoxy-D-xyulose 5-phosphate reductoisomerase	BE095287
Transketolase, chloroplast precursor	AW990934	Tubulin-beta	AW832664	2-cys peroxiredoxin BAS1 precursor	BE095288
Translation Elongation Factor Tu precursor mitochondrial	AW990935	Ubiquitin Conjugating Enzyme E2 17 kDa	AW990943	6-Phosphogluconolactonase	AF222805
Translation Initiation Factor 5A	AF225297	Ubiquitin protein ligase	AW990945	14-3-3	BE095293
Translation Initiation Factor 5A	AW990938	Ubiquitin/ribosomal protein	AW990946	14-3-3	BE095289
Translation Initiation Factor 5A	AW990939	UMP/CMP kinase	BE095280	26S Proteasome AAA-ATPase, regulatory subunit	BE095290
Translationally controlled tumor protein	AW990940	VAMP-associated protein	BE095281		
Translocon-associated protein, beta subunit	AW990941	Vegative storage protein	BE095282		
Transport Protein Sec61, alpha subunit	AW944712	Vegative storage protein	BE095283		
Tubulin-alpha 1 chain	AW990942	WD-40 repeat-protein	BE231356		
		Zinc-finger protein	BE095284		

TABLE 2. Partial characterization of an expressed sequence tag (EST) database for *Euphorbia esula* underground adventitious shoot buds. Designated categories representing 1% or greater of the ESTs.

Designated category	%
Unknown proteins	29.8
Ribosomal proteins	10.0
Cell cycle-associated proteins	3.6
Heat shock or chaperonin-type proteins	3.6
Hormone-regulated or -associated proteins	2.1
Elongation factors	1.8
Antioxidant- and detoxification-associated proteins	1.6
Serine/threonine receptor-like protein kinases	1.6
Signal transduction proteins (putative)	1.6
Chlorophyll a/b-binding proteins	1.2
Aquaporins	1.0
DNA-binding proteins	1.0
Peroxidases	1.0

2000). To determine if these leafy spurge genes are differentially expressed, they were used to probe northern blots of RNA collected from buds at various times after defoliation (Figure 1). RNA from all of the genes showed differential accumulation with the exception of *CAK* and *Scarecrow-like 7*. There was a marked increase over control levels of *Histone H3* and *GASA* RNA between 24 and 36 h that reached maximum detected levels 48 or 72 h after defoliation for the two experimental runs. Because GA is suspected to control both of these genes, it is not surprising that they are coordinately expressed (Aubert et al. 1998; Sauter and Kende 1992). The amount of RNA hybridizing to the *Tubulin* clone did not show a significant increase prior to 48 h (slightly later than *Histone H3* and *GASA*) but reached maximum levels detected in two experimental runs 48 and 72 h after defoliation. Differential expression of these genes in growing buds is consistent with the idea that genes controlling cell division are responsive to the signal transduction pathways that control dormancy.

CAK activity is required for phosphorylation of CDKs and subsequent induction of the cell cycle (Kato et al. 1994). Leafy spurge mRNA for CAK was present at similar levels throughout the time course (Figure 1). In *A. thaliana*, CAK is expressed primarily in actively growing but undifferentiated tissue and is not highly expressed in growing but differentiated tissue such as seedlings and rosettes (Umeda et al. 1998). Additional studies will be needed to determine if leafy spurge CAK has a similar expression pattern. Also, the *Scarecrow-like 7* homolog of leafy spurge does not show substantial differential accumulation. However, this may not be surprising because this gene encodes a protein involved in signal transduction and thus would be required prior to signal perception. Along with consistent levels of ribosomal RNA visualized by imaging (see Materials and Methods), the lack of significant changes in gene expression for these two genes (*CAK* and *Scarecrow-like 7*) over the time course supports the likelihood of consistent RNA loading on the gels used for generation of the northern blots.

The *Lhcb1* homolog was up-regulated in samples containing developing buds (from 3-d defoliated plants) relative to populations of nongrowing buds (samples from control plants) but did not show maximal induction until 72 h after defoliation. Again, this pattern of expression is consistent with what was expected; however, *Lhcb1* RNA level increases

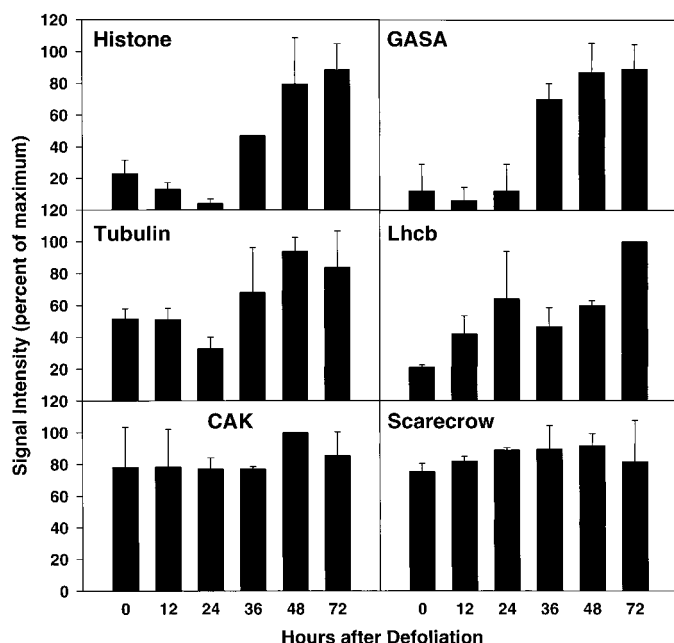


FIGURE 1. Histograms of signal intensity expressed as percentage of maximum \pm standard deviation of northern blots from two separate experiments. RNA was collected from underground adventitious shoot buds at various times (h) following defoliation. Blots were sequentially probed with the indicated genes. *Histone H3* (Accession No. AF239930), *GASA* (GA-responsive gene; Accession No. AW862634), *Tubulin* (Accession No. AW832663), *Lhcb1* (light-harvesting chlorophyll a/b-binding protein; Accession No. AF220527), *CAK* (CDK-activating protein kinase; Accession No. AF230740), and *Scarecrow-like 7* (Accession No. AW944694). Equal loading of RNA for each time point was determined by visual inspection of ethidium bromide-stained ribosomal bands.

occurred earlier than *Histone H3* or *GASA* RNA levels and thus may be controlled by a separate signaling pathway. Recent studies demonstrating that *Lhcb* and *Tubulin* are regulated by auxin, whereas *GASA* and *Histone* genes are regulated by sugars produced in the mature leaves, are consistent with the possibility that the coordinately expressed genes are under the control of similar signals (D. P. Horvath et al., unpublished data).

These data show the immediate effect that can be obtained from the development of an EST database. Not only does our leafy spurge EST database allow us to initiate a genomics-based research program to study the growth and development of buds, but it also provides a genetic resource for other research programs interested in gene expression in perennial plants. Clearly, the development of an EST database for perennial weeds provides a wealth of resources for monitoring the growth and development of vegetative buds at the molecular level. In time, data obtained from *cis*-acting elements associated with differentially expressed genes will help us identify genetic regulatory components that interact with these elements. Future research projects associated with the development of our leafy spurge EST database will enhance our knowledge of the biology of vegetative reproduction in perennials and will provide valuable new information to enhance integrated pest management systems for controlling this noxious weed. This paper is the first of many that will show the potential for developing a genomics-based research program to study weedy characteristics.

Sources of Materials

- ¹ Rootone®, Gulfstream Home and Garden, Lexington, KY.
- ² Sunshine Mix 1, Sun Gro Horticulture Canada Ltd.
- ³ SC-10 Super Cell Ray Leach Cone-tainers®, Stuewe and Sons, Corvallis, OR.
- ⁴ PolyAtract mRNA Isolation System III, Promega Corp., Madison, WI.
- ⁵ Hybri-Zap 2.1 XR two-hybrid vector1, Stratagene, La Jolla, CA.
- ⁶ ABI Prism® BigDye® Terminator Cycle Sequencing Ready Reaction Kit Protocol, PE Applied Biosystems, Foster City, CA.
- ⁷ PTC-200 Peltier Thermal Cycler, MJ Research Inc., Watertown, MA.
- ⁸ DyeEx spin column, Qiagen Inc., Valencia, CA.
- ⁹ GeneQuant spectrophotometer, Pharmacia, Piscataway, NJ.
- ¹⁰ Fluor-S MultiImager, Bio-Rad Corp., Hercules CA.
- ¹¹ Packard Instant Imager, Hewlett-Packard Corp., Palo Alto, CA.
- ¹² Lasergene 99 software program, DNASTAR Inc., Madison, WI.

Acknowledgment

The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the U.S. Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

Literature Cited

- Aharoni, A. L., C. P. Keizer, H. J. Bouwmeester, et al. 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* 12:647–662.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Anderson J. V. and D. P. Horvath. 2000. Isolation of a cDNA clone (Accession No. AF220527) encoding a light-harvesting chlorophyll a/b-binding protein (Lhcb1) in underground adventitious buds of leafy spurge (*Euphorbia esula* L.). *PGR00-052*. *Plant Physiol.* 122:1457.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815.
- Aubert, D., M. Chevillard, A. M. Dorne, G. Arlaud, and M. Herzog. 1998. Expression patterns of GASA genes in *Arabidopsis thaliana*: the GASA4 gene is up-regulated by gibberellins in meristematic regions. *Plant Mol. Biol.* 36:871–883.
- Bangsund, D. A., F. L. Leistritz, and J. A. Leitch. 1999. Assessing economic impacts of biological control of weeds: the case of leafy spurge in the northern great plains of the United States. *J. Environ. Manag.* 56:35–43.
- Biesboer, D. D. and N. Eckardt. 1996. *Euphorbia esula* Element Stewardship Abstract. Arlington, VA: The Nature Conservancy. pp. 1–17.
- Boguski, M. S., T. M. Lowe, and C. M. Tolstoshev. 1993. dbEST database for “expressed sequence tags.” *Nat. Genet.* 4:332–333.
- Chang, S., J. Puryear, and J. Cairney. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11:113–116.
- Chao, W. S., J. V. Anderson, and D. P. Horvath. 2000. Sugar Plays a Role in Inhibition of Underground Adventitious Bud Growth in Leafy Spurge (*Euphorbia esula* L.). Page 46 in *Plant Biology 2000 Final Program and Abstract Supplement*. Rockville, MD: American Society of Plant Biologists.
- Coupland, R. T., G. W. Selleck, and J. F. Alex. 1955. Distribution of vegetative buds on the underground parts of leafy spurge (*Euphorbia esula* L.). *Can. J. Agric. Sci.* 35:161–167.
- Horvath, D. P. 1999. Role of mature leaves in inhibition of root bud growth in *Euphorbia esula*. *Weed Sci.* 47:544–550.
- Horvath, D. P. and J. V. Anderson. 2000. The effect of photosynthesis on underground adventitious shoot bud dormancy/quiescence in leafy spurge (*Euphorbia esula* L.). Pages 30–34 in J.-D. Viemont and J. J.

- Crabbe, eds. 2nd International Symposium on Plant Dormancy: Short Communications. Presses de L'Université d'Angers and CAB International.
- Horvath, D. P. and P. A. Olson. 1998. Cloning and characterization of cold-regulated glycine-rich RNA-binding protein genes from leafy spurge (*Euphorbia esula* L.) and comparison to heterologous genomic clones. *Plant Mol. Biol.* 38:531–538.
- Kamalay, J. C. and R. B. Goldberg. 1980. Regulation of structural gene expression in tobacco. *Cell* 19:935–946.
- Kato, J.-Y., M. Matsuoka, D. K. Strom, and C. J. Sherr. 1994. Regulation of cyclin D-dependent kinase 4 (cdk4) by cdk4-activating kinase. *Mol. Cell. Biol.* 14:2713–2721.
- Keith, C. S., D. O. Hoang, B. M. Barrett, B. Feigelman, M. C. Nelson, H. Thai, and C. Baysdorfer. 1993. Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol.* 101:329–332.
- Kirby, D. R., R. B. Carlson, K. D. Krabbenhoft, D. Mundal, and M. M. Kirby. 2000. Biological control of leafy spurge with introduced flea beetles (*Aphthona* spp.). *J. Range Manag.* 53:305–308.
- Leitch, J. A., F. L. Leistriz, and D. A. Bangsund. 1996. Economic effect of leafy spurge in the upper great plains: methods, models, and results. *Impact Assessment* 14:419–433.
- Newman, T., F. J. de Bruijn, P. Green, et al. 1994. Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* 106:1241–1255.
- Park, Y. S., J. M. Kwak, O.-Y. Kwon, Y. S. Kim, D. S. Lee, M. J. Cho, H. H. Lee, and H. G. Nam. 1993. Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol.* 103:359–370.
- Pennisi, E. 2001. The human genome. *Science* 291:1177–1180.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning, A Laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press. pp. 7.37–7.52.
- Sauter, M. and H. Kende. 1992. Gibberellin-induced growth and regulation of the cell division cycle in deepwater rice. *Planta* 188:362–368.
- Sell, R. S., D. A. Bangsund, L. F. Leistriz, and D. Nudell. 1998. Perceptions of Leafy Spurge by Public Land Managers, Local Decision Makers, and Ranch Operators. Fargo, ND: North Dakota State University, Dept. of Ag. Sci. Agricultural Economic Rep. 406-S. 8 p.
- Stein, B. A. and S. R. Flack, eds. 1996. *The Dirty Dozen. America's Least Wanted: Alien Species Invasions of U.S. Ecosystems*. Arlington, VA: The Nature Conservancy. Available online at <http://www.consci.org/library/pubs/dd/toc.html>. Last accessed July 5, 2001.
- Uchimiya, H., S. Kidou, T. Shimazaki, et al. 1992. Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). *Plant J.* 2:1005–1009.
- Umeda, M., R. P. Bhalerao, J. Schell, H. Uchimiya, and C. Koncz. 1998. A distinct cyclin-dependent kinase-activating kinase of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 95:5021–5026.

Received April 12, 2001, and approved May 18, 2001.